DOCUMENT RESUME

ED 452 231 TM 032 517

AUTHOR Thomas, Paul; Newton, Steve
TITLE Evaluating the Effectiveness of Beginning Teacher Programs in a Large Urban School District.
PUB DATE 2001-04-00
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Beginning Teacher Induction; Beginning Teachers; Elementary Secondary Education; Internship Programs; Program Effectiveness; *Program Evaluation; Teaching Experience; Urban Schools

ABSTRACT

This paper describes the key elements of a model evaluation of beginning teacher programs in a large urban school district. Over a year and a half, evaluators conducted studies of the effectiveness of these programs, some of which were still in progress. This paper reports on the elements used in each study and analyzes the usefulness of each approach. The studies are: (1) 2 surveys asking 581 and 357 teachers about their confidence in the skills identified in the California Standards for the Teaching Profession; (2) an evaluation of the effectiveness of participation in a formative assessment and support program conducted through a telephone survey of 100 participants; (3) the evaluation of the beginning teacher programs as restructured in July 2000 through a study of 30 participants; (4) an evaluation of the special education credentialing program through interviews with 21 teachers, 3 support coaches, and 1 teacher of interns; and (5) an evaluation of the effectiveness of internship programs that will be conducted through evaluation of outcome data. An evaluation of these studies will yield information on the best approaches to evaluating beginning teacher support programs and to improving programs. (Contains 2 figures and 28 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Thomas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 452 231

Evaluating the Effectiveness of Beginning Teacher Programs in A Large Urban School District[1]

Paul Thomas
Steve Newton
Los Angeles Unified School District

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Objective

The purpose of this paper is to identify key elements of a model evaluation of beginning teacher (BT) programs in a large urban school district. The evaluators over the past year and a half have conducted six major studies of the effectiveness of beginning teacher programs in a large urban school district. Some of the studies are still in progress. Each of the studies measured some unique aspect of beginning teacher support. This paper will report on the elements that were used in each study and analyze the usefulness of each.

Review of Research Literature on Teacher Induction

Definition of induction. Simmons (2000) defines teacher induction:

> Teacher induction is the process of socialization to the teaching profession, adjustment to the procedures and mores of a school site and school system, and development of effective instructional and classroom management skills. Participants in these programs are called "inductees," a term which refers simultaneously to teachers who are new to the profession, and teachers with experience who are new to a district, grade level, or certification area (p.2).

Based on the way researchers refer to *induction*, Blair-Larsen, S. (1998) defined teacher induction as "the period of transition from student to professional" (p. 602). Blair-Larsen (1998) cited Huling-Austin (1990) who wrote that teacher induction should be considered as part of the teacher education continuum, that is preservice, induction, and inservice (p. 602).

Lawson (1992) defined *induction* as "the preplanned, structured, and short-term assistance programs offered in schools for beginning teachers" (p. 163). He goes on to state that:

> This new conception owes its popularity to three convergent influences: educational research, political intrusions, and educator's reform proposals. Benefits of this new conception include the attention it gives to beginning

TM032517

[1] Paper presented at the American Educational Research Association Annual Meeting, Seattle, April 2001.

BEST COPY AVAILABLE

> teachers. Its problems include (a) trying to do too much, (b) contriving collegiality, (c) fostering competition among teachers, (d) neglecting teachers' needs, (e) increasing custodial orientations among teachers, and (e) failing to accommodate personal-developmental needs of teachers (p. 163).

Lawson states, however, that "I am not suggesting that all such programs be abandoned. Rather my critique is intended to pave the way for views of teacher induction that retain the advantages associated with this new conception, while addressing its inadequacies" (p. 168).

Essential Components of Induction. Moir and Gless (2001) cite five essential components of a quality induction program: (1) program vision, (2) institutional commitment and support, (3) quality mentoring, (4) professional standards, and (5) classroom-based teacher learning. Klung and Salzman (1991) investigated the effects of two induction models. One was a structured team approach, and the other was a loose buddy system. Findings of the study supported the need for structured induction programs for novice teachers (p. 251).

Much current research points to the importance of attention to key elements of teacher induction. Darling-Hammond (2000), using data from Schools and Staffing Surveys (SASS) and the National Assessment of Educational Progress (NAEP) for 1993-94 and analyses of state policies from all 50 states, found a "strong, significant relationship of teacher quality variables to student achievement even after controlling for student poverty and for student language background" (p.24).

Beginning Teacher Needs. Research continues on the needs of beginning teachers. Darling-Hammond, Berry, Haselkorn, and Fideler (1999) cite studies regarding the fairly minimal support beginning teachers receive (p. 216). This support offers generic strategies for classroom management, but often lacks the subject-specific strategies that work with culturally diverse learners (p. 216). "In turn, many induction programs address the symptoms, not the cause, of disengagement and suboptimal learning, offering a litany of workshops offering formulaic discipline strategies as a form of novice teacher triage" (p. 217).

Many states' induction initiatives focus only on assessment and/or assistance. As a result of states focusing on assessment the support teachers need is often reduced to helping them pass the standardized tests (Darling-Hammond, et al.,1999, p. 217). This support does not always address the needs beginning teachers have to learn to think through challenges to teaching and figure out solutions that address the assessed needs of the students. Regarding assessment as support for beginning teachers, Colbert (1994) pointed out that, "Assessment of beginning teachers may create high levels of stress and anxiety for individuals already faced with a myriad of difficult new tasks and responsibilities" (p. 6).

Support directed solely at financial assistance, while helpful in the short run, does not address the need beginning teachers have for working with a mentor in close supervision and support. Funding for trained mentors is rarely provided by state-mandated induction programs (p. 217).

Darling-Hammond, et al. (1999) cite studies (California Commission on Teacher Credentialing, 1992; Karge and Friberg, 1992) that have found that "well-designed mentoring programs improve retention rates for new teachers along with their attitudes, feelings of efficacy and control, and range of instructional strategies" (p.218). Ingersoll (1997) analyzed a large-scale national teacher survey and found that "where teachers report receiving high-quality assistance, they are also more likely to indicate that they are committed to the profession" (cited in Darling-Hammond, et al., 1999, p. 218).

Induction Partnerships. Some states recognize partnerships between school districts and local schools of education to address teacher induction. Several of these programs were begun in California in the mid-1980s. The California Commission on Teacher Credentialing (1996) issued a report of its evaluation of these programs. Among the conclusions and recommendations were the following:

- The formal assistance by mentor teachers has deteriorated in quality and effectiveness since 1987.
- Demographic information indicates that district intern programs are attracting a high proportion of these who are underrepresented in the teaching workforce.
- Interns in these programs derive a substantial degree of satisfaction from their teaching positions and three-fourths remain in their preparing districts for several years.
- Well over half of the districts that initiated district intern programs terminated the program after training only one or two teachers.
- District intern programs provide an alternative route into teaching for persons whose economic circumstances prevent them from entering teaching through traditional programs.
- Important elements of district intern programs must be improved, such as the unevenness of intern support and the use of District Intern Certificates to provide a convenient hiring mechanism rather than a professional preparation program (pp. 62-64).

Recommendations. Darling-Hammond, et al. (1999) make the following recommendations regarding teacher induction:

- Expand the pool of prospective teachers.
- Reallocate funds from the nonteaching functions that dominate the budgets of American schools to functions that support teacher recruitment and investments in expertise.
- Develop more comprehensive measures of teachers' abilities that sit closer to actual teaching practice and are validated in terms of student learning, such as systematic evidence of practice assembled in highly structured portfolios.

- Provide school districts with greater information about the content and quality of teacher education programs by developing indicators of quality and incentives for universities and schools to pursue them.
- Reduce obstacles to teacher career entry.
- Increase evaluation of various teacher recruitment and prepration initiatives in order to design new program models (pp. 219-224).

They further recommend that "more intensive research regarding teacher recruitment, retention, selection, and induction is needed to inform the policymaking process and the development of successful programs during this critical era of growing teacher hiring" (p. 224).

Research Questions (A Core Set of Research Questions That Cut Across the Individual Studies)

Given the need for teacher induction programs, and hence, the need to evaluate them, and given the need to help guide the developmental stage of induction programs, the following research questions were developed to define the research of this paper. In large urban school districts:

- What strategies are available to study the effectiveness of beginning teacher programs?
- Under what conditions are certain strategies useful in studying the effectiveness of beginning teacher programs?
- What barriers impede studying the effectiveness of beginning teachers?
- What support frameworks for beginning teachers lend themselves to studying effectiveness of beginning teacher programs?

We are particularly interested in presenting the usefulness of studies of effectiveness of these programs. Three key concepts from Patton (1997) will be used to illuminate the differences in the evaluations of teacher induction programs. First, Patton (1997) presents a case for "fostering intended use by intended users" (pp. 41 f.) by taking the following steps:

- Determine who will decide these issues (utilization-focused evaluations).
- Determine what the intended use of the findings will be (judgment oriented, improvement oriented, knowledge oriented).
- Determine the process (intervention oriented, empowerment oriented, development oriented).

Second, Patton (1997) cited Bennett's (1982) conceptualization of the relationship between *chain of events* and *levels of evidence* in program evaluation[2]. The visual of this relationship resembles two sides of a mountain. Each side of the mountain is a scale that has seven matching levels ranging from *inputs* to *end results* on the *chain of events* side and from *resources expended* to *measures of impact* on the *levels of evidence* side. The

[2] See Attachment B.

third concept addresses the relationship between evaluation of process and evaluation of outcome. For programs that are still being developed, there is good reason to put a higher priority on evaluating process than on evaluating outcome. As Patton (1997) explains:

"It is important to know the extent to which a program attains intended outcomes and meets participant needs, but to answer these questions it is essential to know what occurred in the program that can reasonably be connected to outcomes." (p. 197)

Taken together, these concepts indicate that evaluations using multiple measures of process and outcome, though more difficult to implement, may present important advantages for informing decision-makers developing induction programs over more simple evaluations.

A review of the literature regarding theory of teacher induction has been presented. This theory of induction has informed the research questions. Patton's (1997) description of the usability of an evaluation has been briefly presented. The remainder of the paper will describe and qualitatively analyze the six evaluations of beginning teacher induction programs that have either been completed or are in progress. This analysis will provide answers to each of the research questions and draw conclusions about the usability of each of the six evaluations.

Description of Evaluations

The following studies were done in connection with evaluation of effectiveness of beginning teacher induction programs. A visual of these studies together with a conceptualization of the level of utilization of each is presented in Attachment A.

(1) Confidence Surveys #1 and #2. For several years the only evaluation of beginning teacher programs in the District were confidence surveys. The instrument is widely used in California because it is based on the universally recognized California Standards of the Teaching Profession, published by the California Department of Education and the California Commission on Teacher Credentialing. All of the items offered four possible responses. Some comparison between groups of respondents was possible and a pre-/ post- analysis was possible for a portion of the respondents. There were some problems with the survey instrument itself, but because of the broad acceptance of the instrument, the evaluators had no part in selecting or correcting the instrument.

The two surveys were administered in October 1999 and April 2000, respectively. Participants were asked to report their levels of confidence for each of the 31 skills identified in California Standards for the Teaching Profession (CSTP), which is based on the following six standards:

- Engaging and supporting all students in learning
- Creating and maintaining effective environments for student learning
- Understanding and organizing subject matter for student learning
- Planning instruction and designing learning experiences for all students

- Assessing student learning
- Developing as a professional educator[3]

The N was 581 on the first survey and 357 on the second survey. Most of those who took the second survey also took the first survey, so a pre-/post- matched participant gain could be measured. All who participated were beginning teachers participating in either Beginning Teacher Support and Assessment (BTSA) programs or intern programs sponsored either by the District or by local universities.

On both surveys the four point scale options were 1 = very concerned, 2 = a little uneasy, 3 = confident, and 4 = very confident. The mixed anchors of the scale and the limited number of options from which to choose posed problems for the evaluators. Most of the responses (63% in Survey #1 and 83% in Survey #2) were marked either confident or very confident. Interestingly, the alpha for reliability in Survey #1 was .93, and for Survey #2 was .95.

On the first survey one pair of questions indicated an interesting difference in response pattern. The respondents were more confident working with colleagues than in engaging students. Factor analysis on the first survey yielded two distinct groupings in response patterns. No such groupings of response patterns emerged on the second survey. An analysis of matched participant gains between October and April administrations of the survey showed a moderate increase in confidence in key components of CSTP.[4] This comparison indicated that participants showed a moderate increase in confidence since the beginning of their participation in the BTSA program;[5] however, one would have expected some increase in confidence in teaching activities just through practice.

Difficulties in determining usability of the study centered on four key factors. (1) Two different versions of the instrument were printed for Survey #1, which made determination of subgroups impossible; (2) the anchor points were dissimilar on the response scale choices; (3) on Survey #1, 40 respondents did not answer any of the questions on the back of the form; (4) on Survey #2, 24 participants answered all 31 questions with the same response for each item (for example, responding to all items with "3-Confident"); and (5) no control group was used to screen other influences besides participation in the beginning teacher training program (e.g. teaching experience or normal growth by trial and error, without training, over a year). Further, as this paper will demonstrate, the limited functionality of only one outcome measure (levels of confidence) and the use of only one instrument severely diminished the usefulness of the

---

[3] The California Commission on Teacher Credentialing and the California Department of Education and portions by Educational Testing Service (1998).

[4] In order to compare the results of the April 2000 administration of the Confidence Survey with the October 1999 administration, the total scores for each person were calculated and they were compared using a matched pairs t-test (N = 102). This comparison indicated that participants showed a moderate increase in confidence since the beginning of their participation in the BTSA program.[4]

[5] The high degree of statistical significance ($p < .001$, 2-tailed) implies that the difference between the scores was very unlikely to be the result of chance, and the effect size ($d = .639$) indicates that the magnitude of change in confidence since October 1999 was in the medium range (Cohen, 1988).

study to program managers, even though the high levels of confidence might suggest existence of a successful program. The results could only inform program managers to continue more of the same because very little in the findings would have recommended improvements.

A literature review of beginning teacher evaluation studies, done in connection with the Confidence Surveys, revealed that several evaluations have been conducted on BTSA programs, ranging from evaluations focusing on program effectiveness to in-depth evaluations of BTSA programs as a whole. Most of the evaluations have focused on various outcome measures, utilizing self-reports of BTSA participants. The only exception was a 1994 evaluation conducted on individual BTSA programs, the Santa Cruz County BTSA Program in 1994 by Dalton *et al*.

The most comprehensive of the evaluations is the Statewide Evaluation conducted yearly by the California Educational Research Cooperative (CERC). This study evaluates the effectiveness of all BTSA programs in California based upon a standardized assessment instrument, which is completed by beginning teachers, support providers, and site administrators.

The outcome data collected by CERC (Mitchell, Scott, & Boyns, 2000) on the LAUSD BTSA program utilized the Statewide BTSA Evaluation Survey described above but provided data specific to LAUSD. The survey was based on the responses of 387 beginning teachers, 144 support providers, 84 site administrators, and 15 BTSA staff. For each item on the survey, information is presented for each group of BTSA participants, broken down into the categories listed above. For each group, information is presented about the statewide average, LAUSD's deviation from the mean for year 2000 and LAUSD's deviation from the mean from 1999. The interpretation of these scores is left up to the local program that receives the report. The data indicate that beginning teachers rate the BTSA program as lower than the state average on 120 of 134 items. However helpful these scores are for indicating broad problems at that point in LAUSD's program, they do not give any help regarding what can be changed to restore the program to its prior level of success.

A number of important questions that cannot be answered using only the self-report data as collected on the *Statewide Evaluation Survey* (2000) could be answered using other methods of data collection such as interviews with participants, analysis of written documentation, and observation of training and other activities. Important questions that could be addressed through these methods and others include, but are not limited to:

- Are teacher behaviors in the classroom changed as a result of BTSA participation?
- Is teacher retention increased by BTSA participation?
- What factors distinguish successful BTSA programs from unsuccessful ones?
- What are the key characteristics of effective mentors, and how can they be increased?

- How can the tension between assessment and support be navigated so that teachers are supported by their participation in BTSA?
- How are the various contexts identified as important in previous evaluations interacting with the BTSA program to affect various outcome measures?

(2) Evaluating the Effectiveness of Participation in California Formative Assessment and Support System for Teachers (CFASST) for District Interns and BTSA Participants. The next study of beginning teacher (BT) induction support was an evaluation of effectiveness of a framework program whereby beginning teachers and support providers work together in formative assessment techniques to improve the teaching performance of beginning teachers.

A telephone survey of District Intern participants and BTSA participants was conducted during the spring of 2000. A probability sample was drawn from lists of participants provided by the DI/BTSA Office. A total of 100 names were selected (50 first year DIs from a list of 477, and 50 BTSA participants from a list of 1060).

Respondents were asked questions regarding their use of CFASST during the 1999/2000 school year. The CFASST process incorporates the six interrelated categories of teaching practice of CSTP. A major activity of the CFASST program is the Individual Induction Plan (IIP). In consultation with program managers, research questions were developed that focused on the formative assessment activities experienced by BTs, the content of the IIP and the process by which it was developed, evidence of the use of CSTP standards, support activities provided to the BT, and changes in teaching practices brought about by the IIP process. The major findings of the study are the following:

- The degree of beginning teacher CFASST implementation affects beginning teacher outcomes. Respondents who used formative assessment information to guide the development of their IIPs also:
  1) had greater numbers of changes in teaching practice linked to the IIP,
  2) had a greater degree of conceptualization of the relationship between formative assessment and change in the way participants view their teaching, and
  3) demonstrated a greater degree of change in the participant's teaching practice.
- Although 64% of the respondents reported using formative assessment in evaluating their teaching performance, most could not specifically state how formative assessment brought change to their teaching practices or helped them complete elements of the CFASST process.
- The most involved participants in the implementation of the CFASST process were the Year 2 District Interns.

The biggest limitation of this study, hence its limited usability, was a lack of first-hand observation data from beginning teacher classrooms and interviews with support providers. These limitations were addressed with three further studies to be described later in this paper.

With reference to Patton's (1997) exhibit of Bennett's (1982) conceptualization of the relationship between *chain of events* and *levels of evidence*, only data from the first three levels could be gathered with a degree of certainty. The fourth level involving reactions by participants could be measured only to the extent to which the participants had the benefit of completion of the events of CFASST. Their reactions were only as important as their degree of implementation of CFASST.

(3) Evaluation of BTSA Programs Under a Restructured Organization. For several years the BTSA Program has been certified by the State and administered from the LAUSD central office. Beginning in July 2000, however, the District restructured into 11 local districts. Under the restructuring, the BTSA Program is facilitated at the new local district level through a greatly reduced staff at LAUSD central office. Managers are interested in evaluating the delivery of the program as it transitions to the 11 local districts structure. Under the new structure, a cadre of trainers was established in each local district to provide a "Trainer of Trainers" model. Materials are ordered centrally and distributed. Specifically, managers are interested in knowing how local districts will operationalize this model, and articulate their activities with the central district support office in the areas of providing support to beginning teachers, budget and program improvement, and implementation of CFASST.

The following research questions define the instrumentation for gathering information used in this evaluation:

1. How did the delivery of the program change under restructuring?
2. Under what circumstances in the restructured organization did BTSA participants take advantage of the support available to them?
3. How did methods of supervision of the BTSA programs for the District change under the new structure?
4. How was the budget and Program Improvement Plan managed under restructure?
5. To what extent was CFASST implemented?
6. Are there specific ways in which the restructuring program could be adjusted to enhance the BTSA Programs' effectiveness with participants?

Thirty beginning teachers who have voluntarily enrolled in the BTSA program at the local district level have been randomly sampled from three local districts varying in matched average student achievement gains based on SAT/9 tests administered in Spring 1999 and Spring 2000.

Outcome measures include participant, support provider, and program manager reported changes under the new structure; observations of courses and trainings; review of course documents, student outcome measures, and support coach logs; and results of the annual CERC evaluation[6]. Interviews will occur during April and May 2001.

---

[6] California Educational Research Cooperative. University of California, Riverside. An evaluation study done for the Commission on Teacher Credentialing (CTC).

The timeline for the study is September 2000 through August 2001. It is anticipated that this study will provide program managers with more useful information than from previous research because support providers and program managers will be interviewed to enhance the scope of data collected. Further, the use of the CERC evaluations will be compared to the findings of the interviews.

(4) Evaluation of the LAUSD District Intern Special Education Credentialing Program. LAUSD developed the District Intern Program to allow teachers to receive training for a teaching credential while also working in the classroom full-time. It is the only program in the state in which a school district offers such training without the collaboration of a university. Interns commit to a specific track at the beginning of the program, all of which can be completed in two years, except for the Education Specialist track, which takes three years to complete. Program managers initially presented evaluators with a request that they help them to document the success of the Education Specialist program, specifically with reference to its compliance with standards from the California Commission on Teacher Credentialing. Evaluators considered this request, and wondered if it would be the best approach to meet the underlying needs of the program managers. First, no guarantee of a positive evaluation could be made, so the summative evaluation requested by the program managers ran the risk of receiving a negative outcome. Therefore, the evaluators suggested that a more useful approach would be to conduct a formative evaluation of key components and then provide timely feedback so that improvements could be implemented. The program managers accepted and appeared to embrace this suggestion.

Starting in November 2000, interviews with seven year-2 Level I DIs, and 14 year-3 Level 2 DIs were conducted. Interviews with 3 support coaches and one teacher of interns were also conducted. The researcher also observed parts of about six DI courses, and the researcher also observed three trainings for coaches and mentors. The researcher also reviewed documents for courses, portfolios, IIP's, and support coach logs.

Overall DI Specialist Program. The DIs reported facing a variety of challenges in their teaching, many of which revolved around the distinctive obstacles faced by Special Education teachers. DIs noted the following challenges most frequently: the variety of learning abilities in their classrooms; difficulty with curriculum because it was either not provided, or the DI was teaching several different subjects in middle school and therefore had to master curriculum for different subjects; dealing with a lack of respect and support from the local school administration; and, mainstreaming students. DIs reported using a variety of resources to deal with these problems, with most mentioning two or more that were helpful. The most commonly helpful resources reported were: other Special Education teachers at the school site; DI courses; Special Education Mentors; school administrators; School Psychologists and Program Specialists; Support Coaches; and, other DIs. It is interesting to note the congruence between the challenge of teaching Special Education and the helpfulness of people with expertise in Special Education. Also noteworthy is the breadth of support DIs accessed.

When given an opportunity to comment on the program, about one half of the DIs described the program as a whole as "excellent." Two used the metaphor of a "wealth" of resources available to the DIs, and this concept was echoed by several others in different terms. Specifically, DIs with a positive overall impression of the program emphasized support received from various sources such as program administrators, fellow DIs, and formal support providers and the value of the instruction received in the DI classes. About half of the DIs made suggestions for improving the program in response to the same question, with most recommending minor changes to the program. Criticisms included: sometimes unclear assignments, last minute changes of location for classes, and inflexibility regarding classroom attendance. Two of the DIs had stronger criticisms of the program and questioned the quality of the content of the courses, saying they did not get useful information from them.

When asked to recommend improvements to the program, most DIs recommended small changes rather than presenting major criticisms. The most common suggestions focused on modifying the existing program by improving specific components such as the portfolio, or increasing the depth of information taught in the courses, or being more organized.

Other themes regarding the program were specific either to certain program components or to specific DIs. Several Year Two DIs reported that the summer orientation prior to their first teaching in a Special Education classroom was extremely helpful to them. DIs who had prior teaching experience tended to be especially positive about the resources provided by the program. Textbooks for at least some Level II courses have been changed to be more rigorous and research-based. Because of these changes, as well as changes in the tests taught in the advanced assessment course, instructors for some courses have not been able to distribute an accurate syllabus until the second class session.

Formal Support. DI Specialists are assigned both a Special Education Mentor and an Instructional Support Coach. At the time of the interviews, about 80% of DIs reported currently having both a Special Education Mentor and either an Instructional Support Coach or a General Education Mentor, or all three. Of these DIs with more than one formal support provider, 2/3 reported receiving effective support on every issue on which they asked for help. Two DIs complained of having too much support. One DI had no formal support provider at the time of the interview, though the DI had earlier had a Mentor who quit. This DI did not report making an effort to get another Mentor and did not call the Instructional Support Coach that was assigned. Discipline and classroom management were the major areas dealt with in mentoring relationships.

Half of the interns reported getting adequate help in all of the issues they addressed with their mentors. Most of the DIs who did report unresolved issues described situations where circumstances prevented a successful outcome rather than an inability to get good help from the mentor. For example, one DI described how the school administrators refused to authorize a full Behavior Intervention Plan after he had completed a functional analysis. However, two reported not getting adequate help

dealing with significant behavioral issues. A few DIs had problems getting in touch with their off-campus Special Education Mentors, though this difficulty was more of an annoyance than an insurmountable obstacle to receiving support. The most significant obstacle to receiving support was experienced by DIs who were on year-round school calendars. 4 of 5 DIs about whom there was clear data with year-round schedules did not receive formal mentoring until the Fall.

Although DIs were able to get sufficient help on most issues, the Support Coaches and Mentors displayed ignorance about the distinction between Level 1 and Level 2 DIs, and did not utilize the distinction between these levels in their support. One Mentor displayed her confusion in a training by correcting a presenter by saying, "You mean Level 3 (sic)." Mentors and Support Coaches also gave little input on IIP's and Portfolios, which will be discussed later in the report. Training was given by the DI program staff for Mentors and Support Coaches on both the difference between Level I and Level II and also on IIP's and Portfolios after the time they were interviewed.

Assessment. Nineteen of 21 DIs reported using more than one assessment instrument, and the other two both reported only using the KTEA. The most common tests administered were informal tests, followed by the KTEA, and the Brigance. 60% of the DIs administered multiple standardized, normed tests. All but one of the remaining DIs administered one standardized, normed test. For those who administered one or fewer standardized tests, the reason given was that only one test was available at a given school or the students in their classrooms did not need more testing at this time. All DIs reported feeling mostly confident in administration, scoring, and interpretation of the tests. About half of the Level 1 DIs reported minor problems related to inexperience using the tests. Most DIs reported using the results of the tests, especially informal tests, in their classrooms.

There was general satisfaction with the training provided in both the beginning and advanced assessment courses. However, two DIs had very negative opinions of the Level II course. The instructor of the advanced assessment course demonstrated expertise in the tests described, and was able to answer questions from experience and accurately. Training on the Woodcock Johnson III was not possible due to the test being unavailable in time for the course, and so DIs were trained on the Woodcock Johnson – Revised instead. The instruction provided by the psychologist in the Level II assessment course was regarded as very helpful and effective. There appeared to be barely enough time to cover the topics, and perhaps more time, or fewer topics could make the course more effective.

Behavior. Eighty percent of DIs reported using behavioral principles in their classrooms. Only 4 reported doing a Behavior Intervention Plan, but none reported having insufficient knowledge to do one. One DI who did complete a Behavior Intervention Plan reported that the depth of understanding gained was very useful for understanding less extreme behaviors in her classroom, and gave her more confidence in dealing with other behavior issues in the classroom.

Portfolio/IIP. In general, the knowledge of the portfolio and IIP and the support provided to DIs by support providers up to this point of the year on these issues was inconsistent. More than half of the DIs, including Level 2 DIs, had not received help with their IIP or portfolio from a formal support provider, although a few described plans in the near future to do so. Mentors and Support Coaches seemed to have little understanding of the IIP and portfolio. However, it should be noted that the Mentors and Support Coaches received training on IIP's and portfolios immediately after the interviews with DIs and with the Support Coaches. Furthermore, many of the portfolio and IIP components are due in the latter part of the school year, and so can only be evaluated in the next stage of the evaluation.

Conclusion. This report reflects the findings collected in the middle of an evaluation that is to continue for the rest of the year. In general, the DI Specialist Program appears to be providing DIs with resources in a variety of forms that has helped them to deal effectively with the various challenges they have faced as beginning Special Education teachers. Few exceptions to this success were found regarding formal support, assessment, and behavior. The IIP and portfolio process showed inconsistent knowledge on the part of the support providers, but training was given to them regarding these issues, and may change this situation. Further data collection through direct observation of DI classes, observation of DIs utilizing assessment instruments and behavior interventions in their classrooms, and examination of completed portfolios and IIPs will clarify some of the issues before the completion of a final report in July, 2001.

(5) An Evaluation of the Effectiveness of Internship Programs. As a shortage in supply of teachers became evident during the 1980s, alternative certification programs were implemented in California. The internship program allows students to teach while they are completing their course work as long as they have a bachelors degree, are enrolled in either a university intern program or in a district intern program and have passed the California Basic Educational Skills Test (CBEST) and either the Multiple Subjects Aptitude Test (MSAT) or Single Subject Aptitude Test (SSAT).

As a follow-up study to the work that has been done by CTC and others, this evaluation examines effectiveness of internship programs in LAUSD. The purpose of this evaluation is to study the effectiveness of internship programs, not to compare District Intern programs with University Intern programs. However, analytic comparisons will be made between alternative internship programs and traditional certification programs. Research questions for this study are the following:

1. To what extent are interns able to design classroom assignments that generate learning?
2. To what extent are interns able to analyze student achievement on classroom assignments according to clear grading criteria based on teacher goals for the assignment?
3. To what extent do interns demonstrate presentation skills by engaging all learners in critical thinking experiences?

4. To what extent do interns provide students with a classroom environment and climate conducive to learning? What is the level of matched student gains in achievement?
5. To what extent do interns demonstrate investment in their jobs and apply knowledge of how to deal with classroom situations?
6. Under what conditions are interns able to reflect critically on their instructional practices?

Beginning in April and running through May 2001, a probability sample of 30 second year interns working for LAUSD will be interviewed and their classroom instructional practice will be observed. These interns will be matched with 30 second year probationary teachers at their same schools. These comparison-sample teachers will also be interviewed and observed.

In addition, the following data will be gathered: SAT/9 Test matched student gain results, Reading Instruction Competence Assessment (RICA) results, classroom assignments, student portfolio evaluations, intern knowledge application scenario responses, intern retention rates, intern teacher training institution course syllabi, intern training classroom observation, interviews with intern instructors, and CSTP-related assessments done by support providers.

Classroom assignment rating as an indicator of classroom practice has been piloted not only by CRESST/University of California, Los Angeles (see Aschbacher, 1999 and Clare, 2000), but also by the LAUSD Program Evaluation and Research Branch (PERB). PERB is currently using this approach in 55 schools across 11 local districts within LAUSD. It is anticipated that this tool will become a standard rating instrument for teachers at certain specific grade levels in LAUSD by school year 2001-02. Classroom assignments (including elements such as lesson goals, student directions, and grading rubric) along with copies of teacher-graded student work samples are submitted to the evaluation team. These assignments are rated by two independent judges on a four-point scale according to the following seven dimensions of quality:

- Cognitive challenge of the task
- Clarity of the teacher's goals for student learning
- Clarity of the grading criteria
- Alignment of learning goals and task
- Alignment between the goals and grading criteria
- Overall quality of the assignment.
- Fosters critical thinking skills and allows students to demonstrate those skills.

The outcome data for this study will describe the effectiveness of interns practicing in LAUSD. Several measures will assess knowledge, teaching skill, quality of coursework interns take, and performance of students of interns. It is also anticipated that the research design piloted in LAUSD may guide other districts that wish to formally evaluate the effectiveness of internship programs as alternative routes to teacher certification in California.

It is anticipated that this study will provide program managers with the following improvements over previous research:

- Classroom observations and face-to-face interviews will provide data gathered closer to the actual service delivery.
- Findings can be triangulated over several different data collection instruments and procedures.
- Findings will be more useful to program managers by providing a clearer picture of intern instructional practice.

Summary

The perspective of this paper is that of evaluators working within a large urban school district. Although managers of beginning teacher programs may have supported the effectiveness of their work with satisfaction surveys alone in the past, they have come to appreciate a more analytic approach which usually provides more accurate, deeper, relevant, more policy-focused information. Although satisfaction surveys may on the surface yield a flattering evaluation of programs, they give little reliable information about what specific support strategies produce change in instructional methods and why. Reports need to emphasize what a teacher is doing as opposed to how she feels. A more informed investigation needs to provide richer findings by including personal interviews, case studies, observations of actual classrooms and analysis of substantive numerical data.

This paper addresses key questions from managers and evaluators of beginning teacher programs. What are the best evaluation methods for determining the effectiveness of beginning teacher programs? What major issues are easily overlooked? What unique beginning teacher needs should be identified in order to determine what determines BT growth in performance over time? Should beginning teachers be required to participate with their mentors in some framework of formative assessment of their teaching practice (i.e. CFASST)? The best way to address these issues is to determine what evaluation measures yield the most informative results as important decisions are made regarding the development of a large urban school district's most fundamental resource, its teachers.

Studies yielding high reliability and effect sizes may not guarantee that results will be as used by the stakeholders. Studies with smaller numbers of participants done in greater depth may be more meaningful. In the program of studies described above, the confidence surveys had high effect sizes and alphas, but they could not provide the degree of information to project managers as the latter studies where multiple measures are being used. As these studies progress, the analysis will pursue the most useful methods of evaluation which allow for not only gathering evidence for knowledge, attitude, and skill changes, but also for practice and behavior change, including data collection as close to the service delivery as possible (e.g. personal interviews and classroom observations as opposed to surveys only). Further sources of reliable data collection for evaluation of effectiveness of beginning teachers are knowledge application

scenarios, analysis of classroom assignments, reviews of beginning teacher portfolio and formative assessment results, interviews with support providers and college instructors, and comparison with other sources of beginning teacher evaluations.

It is anticipated that these sources of evidence will provide clearer answers to the core set of research questions that cut across the individual studies; however, at this point some preliminary answers are available to the research questions we posed earlier. In large urban school districts:

- What strategies are available to study the effectiveness of beginning teacher programs? This summary of studies has presented several strategies not commonly used in studies of effectiveness of induction, among them: classroom observations, analysis of classroom assignments, analysis of mentor-guided formative assessment measures, interviews with providers of instruction and managers of induction programs, analysis of knowledge application scenarios, and analysis of inductee proficiency test results.
- Under what conditions are certain strategies useful in studying the effectiveness of beginning teacher programs? Surveys are useful as a beginning point for evaluation. They give an overall picture of subjective measures of need, confidence, and satisfaction. They can even provide a desirable picture of a program's activities; however, assure accuracy, measures that are less subjective, such as first hand observation are needed.
- What barriers impede studying the effectiveness of beginning teachers? Evaluators' efforts can be frustrated by improperly constructed instruments and by instruments that do not answer the questions asked by those who need to know; however, evaluators can find important information by staying with a flawed instrument, probing it for information that lack of precision in the mechanics of the instrument may hide. Evaluators can assist program managers in determining the purposes of a study. As an example from this summary of studies, evaluators helped managers frame the focus of the DI Specialist study by shifting from a compliance investigation to an evaluation of effectiveness of program implementation.
- What support frameworks for beginning teachers lend themselves to studying effectiveness of beginning teacher programs? Interviews with support providers and program managers are expected to yield information concerning support effectiveness when those interviews are compared with observations of classroom practice and interviews with the beginning teachers themselves. These comparisons will be particularly interesting if the support providers and beginning teachers are engaged in a process of mutual reflection and formative assessment.

The usability of studies of effectiveness of beginning teacher support programs hinge on two factors: (1) Are the data gathered from participants and support providers relevant to the needs of the program managers so that change, if needed, can occur? Are the data accurate, broad enough in scope to yield reliable evidence, sufficiently deep to reflect what beginning teachers need to grow in their induction process? (2) Are the data gathering methods diverse enough to report every aspect of teacher induction, from the

training and support quality provided the participants to the ability and initiative of the participants? Are the data gathering methods versatile enough to demonstrate with precision the participant levels of achievement being attained?

REFERENCES

Aschbacher, P. (December 1999). *Developing indicators of classroom practice to monitor and support school reform.* (CSE Technical Report 513). Los Angeles, CA: CRESST/University of California, Los Angeles.

Blair-Larsen, S. (1998). Designing a mentoring program. *Education 118* (4), 602-605.

California Commission on Teacher Credentialing (1996). *The effectiveness of district intern programs of alternative teacher certification in California: A longitudinal study.* (ERIC Document Reproduction Service No. ED 437 378).

_____.(1997). *California standards for the teaching profession: A description of professional practice for California teachers.* State of California. Sacramento, CA.

_____. (1997). *Standards of quality and effectiveness for beginning teacher support and assessment programs: A description of professional induction for beginning teachers.* State of California. Sacramento, CA.

_____. (2000). *RICA exam information: Schedule, content specifications, performance characteristics, and scoring scales.* Retrieved January 2, 2001 from World Wide Web: http://www.ctc.ca.gov/profserv/examinfo/ricaexam.html

California Department of Education (1999). *Goals 2000 annual report: Introduction and overview.* Retrieved January 2, 2001 from World Wide Web: http://www.cde.ca.gov/search

Clare, L. (November 2000*). Using teachers' assignments as an indicator of classroom practice.* (CSE Technical Report 532). Los Angeles, CA: CRESST/University of California, Los Angeles.

Colbert, J., Wolff, D, & Trimble, K. (1994). *The relationship between assessment and support of beginning teachers: Where do we go from here?* Paper presented at the American Educational Research Association Annual Meeting, New Orleans, 1994 (ERIC Document Reproduction Service No. ED 368 795).

Dalton, S., & others. (1994). *A program evaluation of support and assessment of beginning teachers: A responsive-teaching alternative to diagnosis-and prescription.* (ERIC Document Reproduction Service No. ED 377 176).

Darling-Hammond, L. (1990). Teaching and knowledge: Policy issues posed by alternate certification for teachers. *Peabody Journal of Education* 67 (3), 123-154.

_____ (1997). *The right to learn.* San Francisco, CA: Jossey-Bass Publishers.

_____ (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives 8* (1). Retrieved February 2, 2001 from World Wide Web: http://epaa.asu.edu/epaa/v8n1.

Darling-Hammond, L., Berry, B., Haselkorn, D., & Fideler, E. (1999). Teacher recruitment, selection, and induction. In L. Darling-Hammond & G. Sykes (Ed.). *Teaching as the learning profession: Handbook of policy and practice* (pp. 183-232). San Francisco CA: Jossey-Bass.

Klug, B. & Salzman, S. (1991). Formal induction vs. informal mentoring: Comparative effects and outcomes. *Teaching and Teacher Education 7*, 241-251.

Kwiatkowski, M. (2000). *Debating alternative teacher certification: A trial by achievement*. Retrieved January 4, 2001 from World Wide Web: http://www.edeexcellence.net/better/tchers/15.htm

Lawson, H. (1992). Beyond the new conception of teacher induction. *Journal of teacher education 43* (3), 163-173.

McKibbin, M. (1988). *Alternative certification in California*. Teacher Education Quarterly 15 (3), 51-59.

_____ (1992). *Alternative routes to teacher certification in California: A report to the Legislature*. (Report to the Legislature, June 5, 1992). Sacramento, CA: California Commission on Teacher Credentialing.

_____ (2001). One size does not fit all: Reflections on alternative routes to teacher preparation in California. *Teacher Education Quarterly* 28 (1), 133-149.

Mitchell, D., Scott, L., & Boyns, D. (2000, June). *California beginning teacher support and assessment program: 2000 evaluation survey report for LAUSD Los Angeles County district intern program*. California Educational Research Cooperative (CERC), University of California, Riverside. Riverside, CA.

Mitchell, D., Scott, L., Hendrick, I., & Boyns, D. (1998, June). *California beginning teacher support and assessment program: 1998 evaluation survey report for LAUSD Los Angeles County district intern program*. California Educational Research Cooperative (CERC), University of California, Riverside. Riverside, CA.

Moir, E., Gless, J. (2001). Quality induction: An investment in teachers. *Teacher Education Quarterly 28* (1), 109-114.

Newton, S. (2000). *District intern/BTSA program: Confidence Survey Report*. Los Angeles, CA: Program Evaluation and Research Branch (PERB), Los Angeles Unified School District (LAUSD).

Patton, M. (1997). *Utilization-focused evaluation*. Thousand Oaks CA: Sage Publications.

Simmons, A. (2000). An overview of teacher induction: Questions & answers. *Future Teacher* 7 (1), 2,3. (A newsletter of Recruiting New Teachers, Inc., Belmont, MA).

Thomas, P. (2000). *District intern/BTSA program: Confidence Survey Report*. Los Angeles, CA: Program Evaluation and Research Branch (PERB), Los Angeles Unified School District (LAUSD).

Tyson, S., Hawley, H., McKibbin, M. (2000). *Pre-internship teaching program: A progress report to the legislature*. (Report to the Legislature, October, 2000). Sacramento, CA: California Commission on Teacher Credentialing.

Teacher Training Evaluations

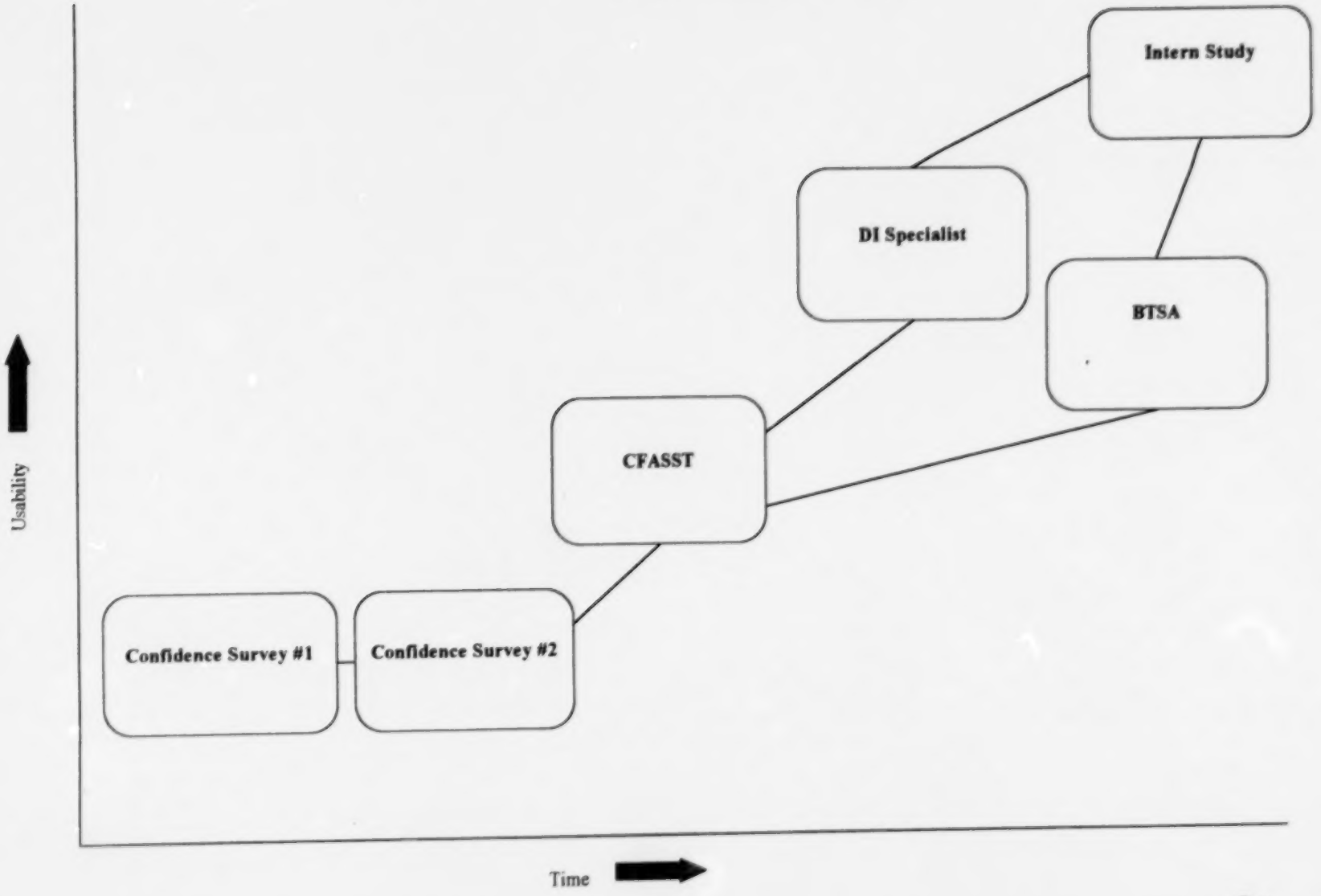
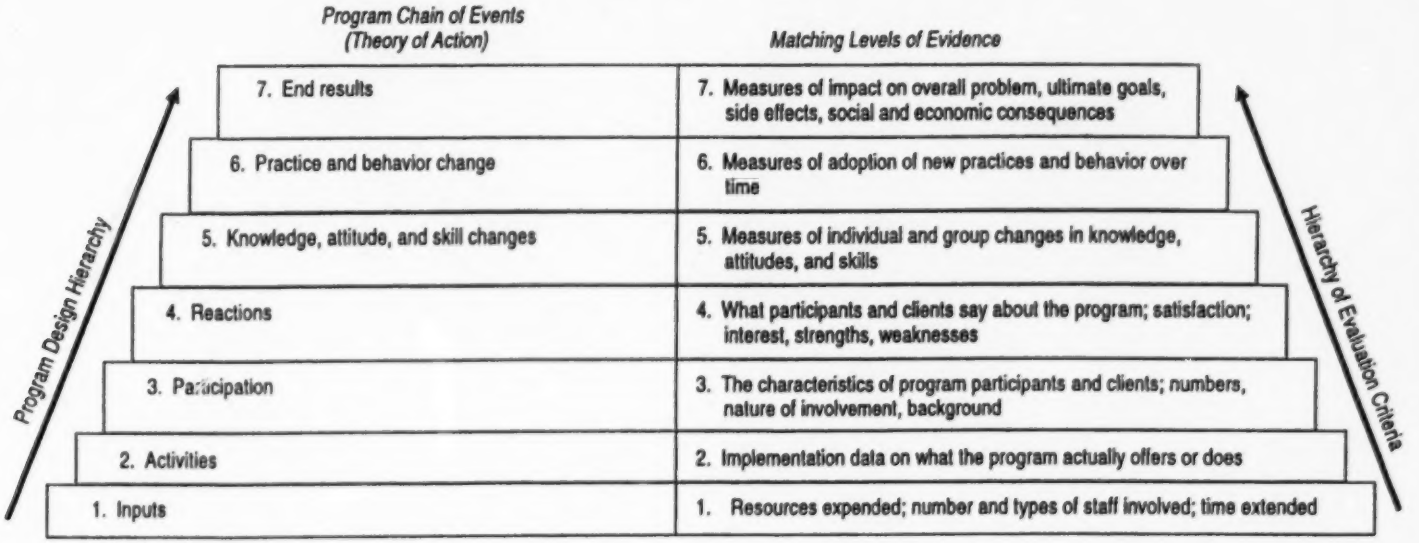


EXHIBIT 10.5


Program Chain of Events
(Theory of Action)
Matching Levels of Evidence
7. End results
7. Measures of impact on overall problem, ultimate goals, side effects, social and economic consequences
6. Practice and behavior change
6. Measures of adoption of new practices and behavior over time
5. Knowledge, attitude, and skill changes
5. Measures of individual and group changes in knowledge, attitudes, and skills
4. Reactions
4. What participants and clients say about the program; satisfaction; interest, strengths, weaknesses
3. Participation
3. The characteristics of program participants and clients; numbers, nature of involvement, background
2. Activities
2. Implementation data on what the program actually offers or does
1. Inputs
1. Resources expended; number and types of staff involved; time extended
Program Design Hierarchy
Hierarchy of Evaluation Criteria

SOURCE: Adapted from Bennett 1979.